

Analyzing γ rays of the Galactic Center with Deep Learning

Sascha Caron^{a,b} Germán A. Gómez-Vargas^c Luc Hendriks^a
Roberto Ruiz de Austri^d

^aInstitute for Mathematics, Astrophysics and Particle Physics, Faculty of Science, Mailbox 79, Radboud University Nijmegen, P.O. Box 9010, NL-6500 GL Nijmegen, The Netherlands

^bNikhef, Science Park, Amsterdam, The Netherlands

^cInstituto de Astrofísica, Pontificia Universidad Católica de Chile, Avenida Vicuña Mackenna 4860, Santiago, Chile

^dInstituto de Física Corpuscular, IFIC-UV/CSIC, Valencia, Spain

E-mail: scaron@cern.ch, ggomezv@uc.cl, luc@luchendriks.com, r Ruiz@ific.uv.es

Abstract.

We present the application of convolutional neural networks to a particular problem in gamma ray astronomy. Explicitly, we use this method to investigate the origin of an excess emission of GeV γ rays in the direction of the Galactic Center, reported by several groups by analyzing *Fermi*-LAT data. Interpretations of this excess include γ rays created by the annihilation of dark matter particles and γ rays originating from a collection of unresolved point sources, such as millisecond pulsars. We train and test convolutional neural networks with simulated *Fermi*-LAT images based on point and diffuse emission models of the Galactic Center tuned to measured γ ray data. Our new method allows precise measurements of the contribution and properties of an unresolved population of γ ray point sources in the interstellar diffuse emission model. The current model predicts the fraction of unresolved point sources with an error of up to 10% and this is expected to decrease with future work.

Contents

1	Introduction	2
2	Convolutional Neural Networks	4
3	Images for training and testing the ConvNets	4
3.1	Template fitting procedure	5
3.2	Simulation of <i>Fermi</i> -LAT data	8
4	Neural network setup	11
4.1	Network architecture	12
5	Results	14
5.1	Neural network results	14
6	Conclusion	17
	Appendices	18
A	ConvNets	18
A.1	ConvNet data pipeline	18
A.1.1	Layer Architectures	19
A.1.2	Choice of activation function	20
A.2	Training a neural network	21

1 Introduction

The *Fermi* Large Area Telescope (LAT) has been observing the γ ray sky for energies $\gtrsim 100$ MeV since August 2008 with unprecedented angular resolution and sensitivity [1]. The γ rays produced by the interaction of cosmic ray (CR) particles with matter and radiation fields in the interstellar medium (ISM) of the Milky Way dominate the emission detected with the *Fermi*-LAT in the Galactic plane [2]. Physical modeling codes, such as GALPROP¹, calculate the propagation of CRs in the ISM and compute diffuse γ ray emission in the same framework [3, 4]. Each GALPROP run using specific, realistic astrophysical inputs will create a different interstellar emission model (IEM) in the form of a series of templates, each for a particular physical emission mechanism. The usual method to compare these IEMs with data is the so-called *template fitting method* [5]. A large set of IEMs were compared to *Fermi*-LAT data in [6], finding that the IEMs reproduce general features of the interstellar γ ray emission over the whole sky. However, the template models do not capture all the information contained in the *Fermi*-LAT data. The present analysis focuses on an unaccounted component detected in the region of the inner Galaxy, the so-called Galactic center Excess (GCE), using the template fitting method [7–16].

The *Fermi*-LAT collaboration investigated the GCE properties (spectral shape, magnitude, and morphology) by using 6.5 years of observations in Refs. [17, 18], with the conclusion that the GCE is present in the data despite many sources of uncertainties, but its specific properties are significantly model dependent. The gray band of figure 1 presents the set of spectral energy distributions of the GCE found in [18], derived from the different systematics sources investigated. This emission has been interpreted as a possible signal for the annihilation of dark matter (DM) particles or as a signal of unresolved point sources [15, 19–21]. DM-like signatures observed in other regions of the Galactic plane, where such signals are not expected (control regions), are used to quantify the magnitude of systematic uncertainties due to diffuse emission modeling in the Galactic center and prevent the interpretation of the GCE as a signal of DM [18].

Many groups have investigated the possibility that a population of unresolved γ ray pulsars could be the origin for the GCE emission, see for instance [20, 22–27]. The Galactic bulge is expected to host a population of pulsars due to the high level of past star formation activity in that region [28]. In the whole sky, more than 200 pulsars have been identified in the γ ray band². Furthermore, using novel statistical methods the authors of [29], [30] and [31] claim evidence for the existence of an unresolved population of γ ray sources in the inner 20° of the Galaxy with spatial distribution and collective flux compatible with the GCE. Recently the *Fermi*-LAT collaboration also investigated the pulsar interpretation of the GCE [32]. Performing a new point source search using 7.5 years of Pass 8 *Fermi*-LAT data in a $40^\circ \times 40^\circ$ box around the Galactic center, they confirm the findings of [29] and [30]. However, some arguments [33–37] have been raised against this interpretation of the GCE. These works point out that a hypothetical γ ray millisecond pulsar (MSP) population in the Galactic bulge with similar properties of extant populations, as in globular clusters and the local Galactic disc, is not able to reproduce the total GCE emission. However, these argument are based on the assumption that an MSP population in the Galactic bulge share similarities with MSP populations elsewhere [38]. Because there are no MSPs detected in the Galactic bulge yet, this assumption might not be true. Therefore, the debate on the nature of the GCE is not yet

¹<http://galprop.stanford.edu>

²<https://confluence.slac.stanford.edu/x/5Jl6Bg>

closed, leaving the possibility of having all or a fraction of the GCE due to DM annihilation or other diffuse sources.

The image texture of the GCE can help to unveil its nature; to determine if the GCE is genuinely diffuse or has a granular morphology [29–31]. A completely diffuse GCE would favor the DM interpretation, whereas cumulative emission of point sources too weak to be detected individually would produce a granular morphology³. The *template fitting* approach is not sensitive to this difference. This work aims to present a new method based on convolutional neural networks (ConvNets) to determine whether the morphology of the GCE is diffuse or granular. We apply our approach to the data collected with the *Fermi*-LAT in the region of the Galactic center, and we use simulations of this area to train and validate the network. The present paper is a proof-of-principle work, and we simplify the models for simulations to a level in which the results on the GCE nature provide valuable information for a more sophisticated implementation of the method in a future publication.

We organize the paper as follows: We start with an introduction to the GCE. Section 2 presents the basics on ConvNets. This technology needs large amounts of data to work, and we use realistic simulations for both training and testing the networks; section 3 expands on the setup to create these images. Section 4 presents the ConvNet designed to make predictions on the fraction of granularity present in the GCE, i.e., how much of the GCE is due to a population of unresolved point sources or a truly diffuse source. Results are presented in section 5, conclusions and foreseen applications are discussed in Section 6.

³If the point sources are too dim or too close to each other, it is tough to distinguish them from diffuse radiation. Also, Ref. [39] argues that if some fraction of the DM has dissipative interactions, it can form dense DM clumps resulting in a population of γ ray point sources in the Galactic center region.

2 Convolutional Neural Networks

ConvNets are a class of deep neural networks (DNNs) that are designed to make predictions on visual data [40]. Neural networks are used in many areas of research to solve classification or regression queries and are well suited for high dimensional problems (for example Ref. [41]). The authors of ref. [42] employed a similar methodology for analyzing gravitational lensing. In general, a ConvNet takes an N-dimensional input, transforms it using many linear and nonlinear operations, and produces an M-dimensional output (for example, the information introduced can be an image and the output a probability that the image belongs to a particular class) [43]. The result of a ConvNet can be a prediction that a specific input belongs to a specific type (e.g., object detection) or a projection of a regression problem (e.g., this work). The input of a computer vision task is typically an image; a (w, h, c) -tensor can be a way to represent the data. The w and h in the tensor denote the width and height of the network and c the number of channels, that can be a color filter or energy band. In this work the number of channels is 1: A single energy-integrated counts map of the Galactic Center region, with a specific fraction of the GCE due to a population of unresolved point sources, free to vary between 0 and 1. We call this fraction the f_{src} parameter. $f_{\text{src}} = 0$ means that the GCE is composed of only a diffuse source and $f_{\text{src}} = 1$ implies that the GCE is formed only of the cumulative emission of point sources too dim to be detected individually.

In this work, the ConvNet is trained to predict f_{src} from Galactic center images created with counts in the energy band 1-6 GeV, as this is the band where the GCE is most relevant, see figure 1 [17, 18]. In Appendix A we provide a more thorough introduction to ConvNets, in particular training them (Appendix A.2).

3 Images for training and testing the ConvNets

Typically a large dataset is needed to train a ConvNet because they can have up to millions (or billions) of tunable weights (see Appendix A). To have an extensive training dataset is a challenge in applying ConvNet technology to the *Fermi*-LAT Galactic center data, as there is only one γ ray image of this region (e.g., compared to many different pictures of galaxies for galaxy classification). However, the advanced knowledge of the *Fermi*-LAT instrument [1] and the availability of state-of-the-art interstellar diffuse models [6] make it possible to create realistic mock *Fermi*-LAT data for training and testing examples. The goal of the trained ConvNet is to predict f_{src} regardless of flux and location of the unresolved point sources as well as the IEM. To achieve this goal, we create a set of *1.2 million mock training images* of the Galactic center region based on models fitted to real data⁴. To determine the optimal amount of pictures in the training set, we train a ConvNet multiple times using different training dataset sizes and compare the accuracy of the f_{src} predictions.

The IEM is a crucial component to generate different realizations of the *Fermi*-LAT observation of the Galactic center region. As there is no unique model, we must implement different IEM versions. We use five of the eight alternatives IEMs created to study the systematic effects on supernova remnant (SNR) properties caused by the modeling of the diffuse emission around them [44]. The models selected represent the largest variation in the input parameters used the set of eight IEMs. The starting point is to create models using GALPROP, with a variation on the three input parameters that according to Ref. [2] are the

⁴ For training; we create 1.000 realizations of 1.200 different values of f_{src} (1.2 million images). Three samples of 20.000 other f_{src} values were used to test the ConvNets (60.000 images).

most relevant in modeling the Galactic plane diffuse emission⁵. The output of GALPROP is in the form of separated templates for emission associated with gas (atomic and molecular) in four Galactocentric rings and a single model for the inverse Compton emission. The second step in the creation of the IEMs is to tune individual scaling factors for these different template components using all-sky likelihood fit to *Fermi*-LAT data. Finally, we re-fit the five IEMs selected leaving free only the scaling factors of the gas related emission in the first innermost rings. In table 1 we present our selection and use of the five IEMs in this work. We decided to use the models created in reference [44] as they have more freedom in fitting the diffuse emission in the Galactic plane than other models, as the standard IEM of the *Fermi*-LAT team for point-source analysis [45]. We also included in the model the sources listed in the 3FGL catalog [46]. Finally, we add to the model of the Galactic center area the integrated γ ray flux at Earth, ϕ_s , expected from dark matter annihilation in density distribution, $\rho(r)$, given by

$$\phi_s(\Delta\Omega) = \underbrace{\frac{1}{4\pi} \frac{\langle\sigma v\rangle}{2m_{DM}^2} \int_{E_{\min}}^{E_{\max}} \frac{dN_\gamma}{dE_\gamma} dE_\gamma}_{\Phi_{PP}} \times \underbrace{\int_{\Delta\Omega} \left\{ \int_{\text{l.o.s.}} \rho^2(r) dl \right\} d\Omega'}_{\text{J-factor}}. \quad (3.1)$$

Here, the Φ_{PP} term depends on the particle physics properties of dark matter—i.e., the thermally-averaged annihilation cross section, $\langle\sigma v\rangle$, the particle mass, m_{DM} , and the differential γ ray yield per annihilation, dN_γ/dE_γ , integrated over the experimental energy range from E_{\min} to E_{\max} . The J – factor is the line-of-sight integral through the dark matter distribution integrated over a solid angle, $\Delta\Omega$. Qualitatively, the J – factor encapsulates the spatial distribution of the dark matter signal, while Φ_{PP} sets its spectral character. Note that we assume annihilation of Majorana particles.

We use the gNFW density distribution to model the spatial DM distribution [47, 48]:

$$\rho(r) = \rho_s \frac{r_s^3}{r^\gamma (1 + r_s)^{3-\gamma}}. \quad (3.2)$$

Where r_s is the scale radius (20 kpc) and ρ_s a scale density fixed by the requirement that the local DM density at Galactocentric radius of 8.5 kpc is 0.4 GeV cm^{-3} . To model Φ_{PP} , we follow the DM interpretation of the GCE considered in ref. [14], and use a $m_{DM} = 50 \text{ GeV}$, and the differential γ ray yield from WIMP annihilating into $b\bar{b}$ final state.

3.1 Template fitting procedure

In this work, we use the *Fermi*-LAT data collected between 2008 August 4 and 2015 August 2 (*Fermi* Mission Elapse Time 239559568 s - 460166404 s). To avoid bias due to residual backgrounds in all-sky or large scale analysis, we select the events belonging to the Pass 8 UltraCleanVeto class and use the corresponding P8R2_ULTRACLEANVETO_V6 instrument response functions; they have the highest purity against contamination from charged particles misclassified as γ rays. Furthermore, to reduce the contamination of γ rays from interactions of CRs in the upper Earth's atmosphere, we select events with a maximum zenith angle of 100° . We bin the events into 21 energy bins from 500 MeV to 100 GeV for fitting the IEMs to the real γ ray sky (note that for the ConvNets we use a single energy bin from 1 to 6 GeV). We only use the events converted in the front part of the *Fermi*-LAT, as this is a

⁵CR source distribution, the height of the CR propagation halo, and spin temperature of the molecular gas.

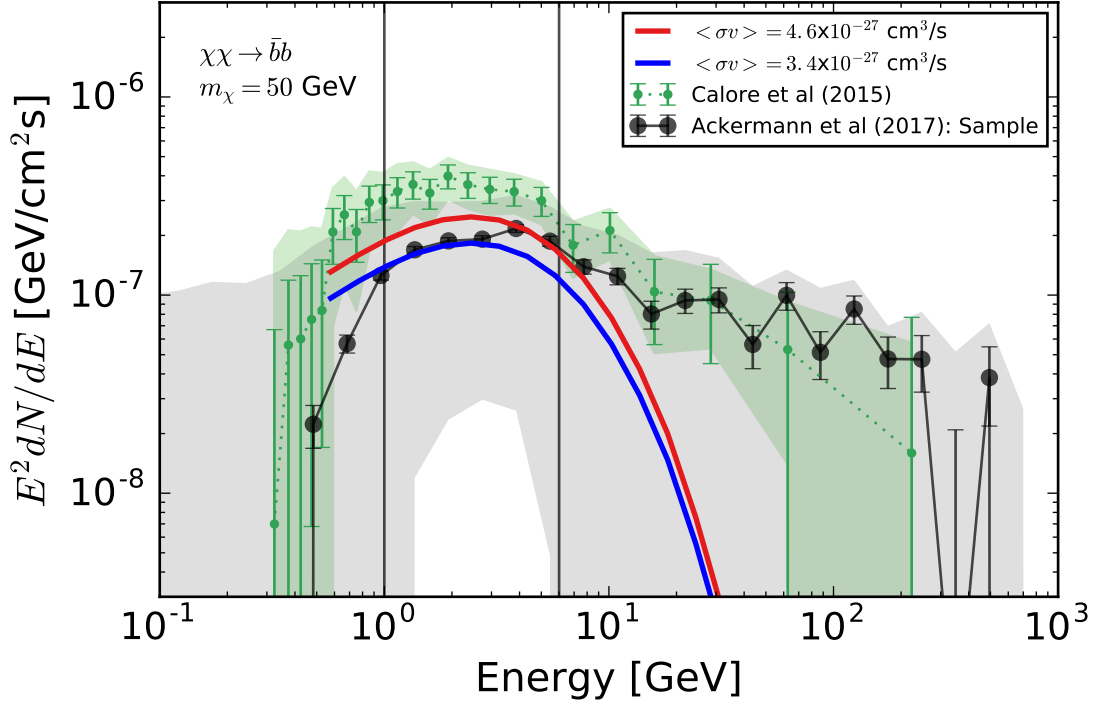


Figure 1: Spectral energy distribution of the GCE. In blue and red the spectra from fits performed in this work to the *Fermi*-LAT photon data using the models Training A and B in table 1, respectively. Following [18] we select the GCE derived using the Sample Model as representative of the GCE spectrum (black points). The gray shaded area represents the variation in the GCE spectra obtained in [18]. The vertical black lines point out the energy range where the GCE is always present; we use this band to create all ConvNets analyses images. For comparison the GCE spectrum as found in [14] (green points) is plotted together with the diagonal of the covariance matrix derived there (light green band).

good compromise between angular resolution and statistics to produce sharp images. Thus we use the same data in fitting and for ConvNets. For the fitting procedure we use a spatial resolution of 0.1° , while for ConvNets we use 0.25° , this is to reduce the Poissonian noise and the computing time of generating images. We re-fit the five IEMs listed in table 1, including a gNFW template, to *Fermi*-LAT data in the inner $15^\circ \times 15^\circ$ about the Galactic center.

Standard *Fermi* tools are used to prepare data and to perform the fits. In this analysis we fit the following set of parameters:

- The normalization of the inner rings in the IEM.
- The normalization of the brightest 3FGL sources (we neither vary the position nor the spectral shape of the point sources, as this would increase the freedom significantly in the fitting). It is worth mentioning that we use a dataset derived with improved reconstruction algorithms and with more years of exposure than in the derivation of the 3FGL catalog.

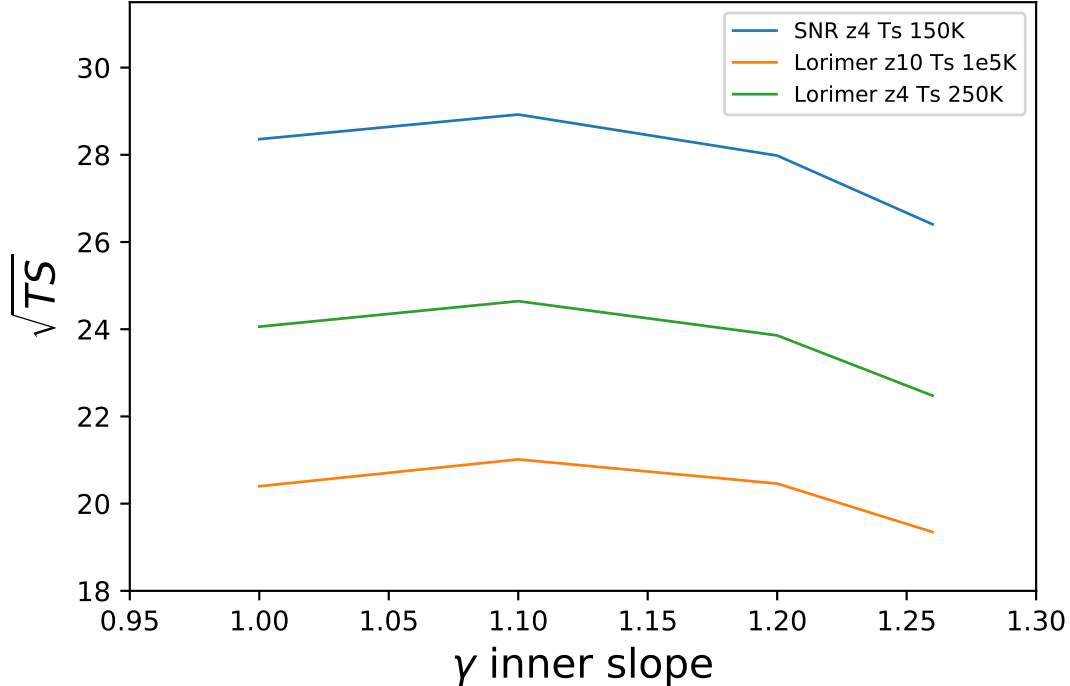


Figure 2: Best fit values for γ using the three different IEMs for training the ConvNet. TS stands for Test Statistics and is defined as $TS = -2\ln\Delta\mathcal{L}$, where $\Delta\mathcal{L}$ is the likelihood ratio between models with and without source, in our case the gNFW component [49]. The inclusion of a gNFW component always improves the quality of the fit. In this proof-of-concept work $\gamma = 1.1$ as this value provides the larger Test Statistics (TS) over the different models tested. It is worth noting that TS is not well defined for extended sources, we use this quantity as an approximated proxy for comparing models, but not to choose the best model, just to select one of them.

- We model the GCE spatially with a gNFW, spectrally as a WIMP of 50 GeV annihilating into $b\bar{b}$ quarks. In this way the only free parameter for this component is the normalization, that we parametrize as the thermally averaged cross section $\langle\sigma v\rangle$.

The internal slope of the gNFW template is varied, changing the γ parameter in eq 3.2. For each value of γ , we run five different fits corresponding to the five IEMs selected. Inspired by the results of [18] the following values for γ are used: 1.0, 1.1, 1.2, 1.26, and 1.3. In the simulations used for training and validating the ConvNets, γ is set at 1.1 as we find a slightly better fit with this configuration, see figure 2. Table 1 presents the likelihood of the five selected IEMs together with the resulting $\langle\sigma v\rangle$ when the IEMs to generate the training and validation data uses gNFW with $\gamma = 1.1$. It is worth noting that all IEMs were previously fitted to all-sky data and based on different assumptions about the initial parameter. Therefore the likelihood ratio test can not be used to compare models as they are not special cases of a null model. However, after inspection of spatial and spectral residuals, we get enough agreement to the data for this methodology analysis.

Table 1: Results of fitting the $15^\circ \times 15^\circ$ region around the GC using five different IEMs. The gNFW template is created with $\gamma = 1.1$. All likelihoods and both spatial and spectral residuals are at the same level, implying that all models provide a similar description of the data. We assume the sources of CRs are supernova remnants (SNRs) considering two CR source distributions, one traced by the observed distribution of pulsars, Lorimer [50], and other tracing SNRs observed [51]. T_S stands for spin temperature of the atomic hydrogen for the derivation of gas column densities from the 21-cm line data.

Usage	CR distribution	Halo height z (kpc)	T_S (K)	$\text{Log}\mathcal{L}$	$\langle\sigma v\rangle \times 10^{-27} \text{ cm}^3/\text{s}$
Training A	SNR	10	150	-442855	45.59
Training B	Lorimer	10	1×10^5	-442304	33.61
Training C	Lorimer	4	150	-442357	39.32
Testing A	Lorimer	10	150	-442539	39.63
Testing B	SNR	4	1×10^5	-442664	42.67

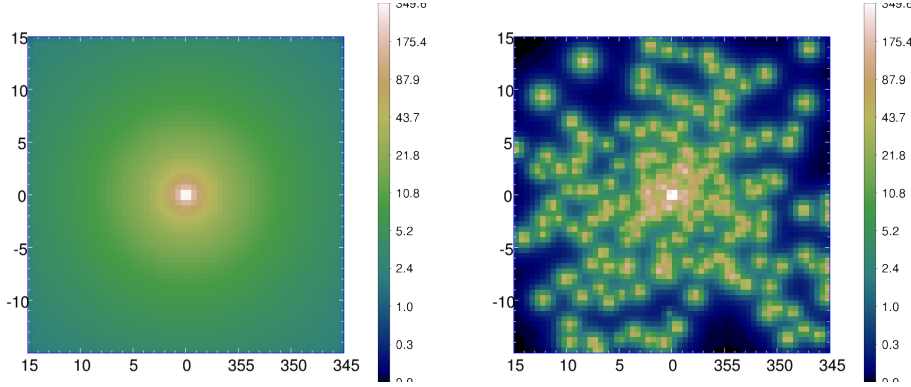


Figure 3: Count maps in the 1-6 GeV energy range comparing the granular vs diffuse nature of the GCE, these templates are indistinguishable for the template method (see section 3.1). Both maps have the same total emission and follow a gNFW distribution.

3.2 Simulation of *Fermi*-LAT data

After fitting the five models of the Galactic Center region described above to the *Fermi*-LAT observations, we generate simulated data using the *gtsrcmap* and *gtmodel* codes of the *Fermi* tools, the former for the convolution of models with the *Fermi*-LAT response, the latter to combine diffuse templates and generate the point source populations. We add Poisson noise to the final images. In this set of fabricated images, we modify the fraction f_{src} between cumulative emission from point sources and a diffuse component of the GCE model, keeping its magnitude, spectral shape and distribution as determined by the fits, see figure 3.

The spectral shape of the GCE is highly dependent on the assumptions of the underlying model, but as can be noted in Figure 1 the excess is always present in the 1-6 GeV band. We, therefore, choose that band to generate images for training and testing the ConvNets. The left panel of figure 3 shows an image made with the gNFW template ($\gamma = 1.1$) in the

1-6 GeV band. The right-hand side of figure 3 shows the image of a simulated population of point sources that produce the same amount of total flux as the gNFW template. The density of sources in the population is also spatially distributed following equation 3.2 with $\gamma = 1.1$. To generate the maps of point source populations which mimic the GCE, we make the following assumptions: all sources have the same spectral shape and are distributed randomly following a density distribution specified above (see figure 4 for a sample plot of a randomized population of point sources). Besides following equation 3.2, we assume a power-law behavior for the flux distribution in the population,

$$\frac{dN_{src}}{ds} = As^\alpha. \quad (3.3)$$

The normalization A is determined by f_{src} times the total emission of the GCE as obtained in the corresponding fit⁶ and the constraint that all sources are below the 3FGL detection threshold ($1 \times 10^{-9} \text{ cm}^{-2} \text{ s}^{-1}$ for fluxes above 1 GeV [46]). s represents the source flux and $\frac{dN_{src}}{ds}$ the number of sources for a specific flux interval. The spectral index α determines the number of point sources per flux below the 3FGL detection threshold. Thus, a high value of α means most point sources are just below to the 3FGL detection threshold, while a low value of α means most point sources are far below the 3FGL detection threshold. For example, a value of $\alpha = 1.05$ implies that almost all sources have a flux close to the detection threshold. In this analysis we let α range from -1.05 to 1.05⁷.

In figure 5 we present some of the models for training, focusing on the extreme cases where the GCE is entirely made by DM annihilation or by a point source population. We generate training and testing data as follows: first, we convolve the best-fit model with the instrumental response function of the *Fermi* LAT (using the *gtmodel* tool of the *Fermi* science tools⁸) yielding sky images with the expected number of counts. Then we generate Poisson instances of these pictures.

One of the main concerns of this approach is the quality of the simulations versus the actual *Fermi*-LAT image of the Galactic center. If the simulations are not representative of the real data, the output is unreliable even if the ConvNet’s accuracy is 100%. The ‘Reality Gap’ is the name of this discrepancy between simulated and real data. We use for simulations state-of-the-art modeling of the diffuse emission to represents the actual data. However, some model parameters that serve as input for the simulation are not well determined by our best-fit procedure. These include:

- Which IEM is correct? The computation of IEMs requires input data that are highly uncertain, the models of our Galaxy represent quite well the whole γ ray sky, but in the inner region, the complexity of the environment is significantly higher [6]. In particular, GALPROP-generated templates are ‘smoothed’ versions of reality as many smaller molecular clouds are not present in the GALPROP-input gas maps. For example, the small-scale structure of the IEM is not present in the training set, see Supplemental Material, item F, of [30]. Furthermore, the models in reference [44] do not enclose the complete range of systematics involved in the modeling of Galactic diffuse emission.

⁶ $\int_s As^\alpha ds$ equals to f_{src} times total GCE flux. The GCE flux varies with the background used in the fit, therefore depending on the background and fraction f_{src} we determine the parameter A .

⁷In [32] the best-fit α value found for the Bulge pulsar population is -1.2. We do not extend our α range to cover that value as we only work with simulations, follow-up work will be devoted to extracting physical quantities from actual *Fermi*-LAT images.

⁸<https://fermi.gsfc.nasa.gov/ssc/data/analysis/scitools/overview.html>

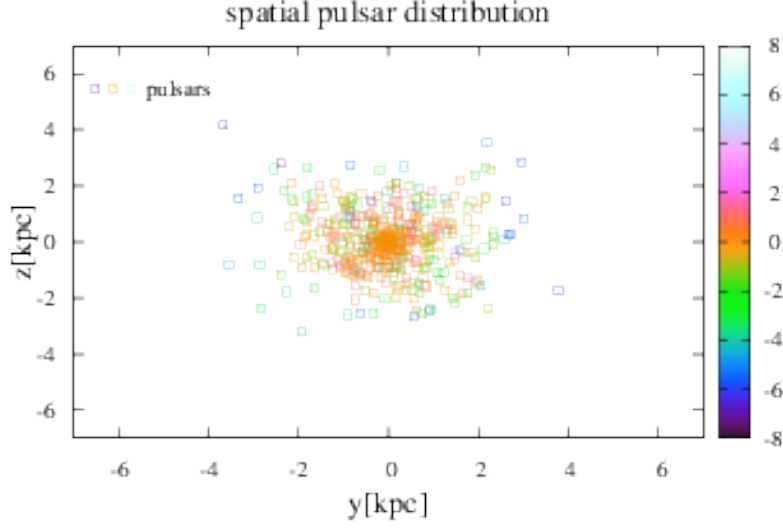


Figure 4: 3D distribution of a point source distribution following an gNFW profile with $\gamma = 1.1$. Each time such a population is generated the sources are placed at different positions. The flux s of each source is also randomly assigned. The (0,0,0) point is the GC, the color code stands for x axis in kpc.

They only consider one method for creating the IEMs and varied three input parameters. For a ConvNet to give accurate results on the real data, the variations of the input parameters need to be large enough so that it can learn to generalize over those parameters.

- The list of detected point sources: we used the 3FGL catalog, which was created using four years of data, while we use about seven years of data to generate the simulations. As new sources are expected to appear with more data, the results of the network are going to be biased towards higher f_{src} . Therefore, sources that could be detected with seven-year data, but not present in the four-year catalog, will in the simulations be part of the signal while they should be included in the background. This miss modeling leads to a higher predicted f_{src} value by the network. The 2FIG catalog presents about 200 more sources detected in the GC region than the 3FGL catalog and used seven years of data, which will remove this bias. For this initial analysis we only wanted to create images similar to the GC but in a simple way, newer catalogs will be part of the models in a follow-up study.
- The positions of point sources part of the unresolved population: we fix only their distribution in the current simulation.
- The true distribution of the GCE emission: in our work we use gNFW template parametrized by γ in equation 3.2, which is here fixed to 1.1.

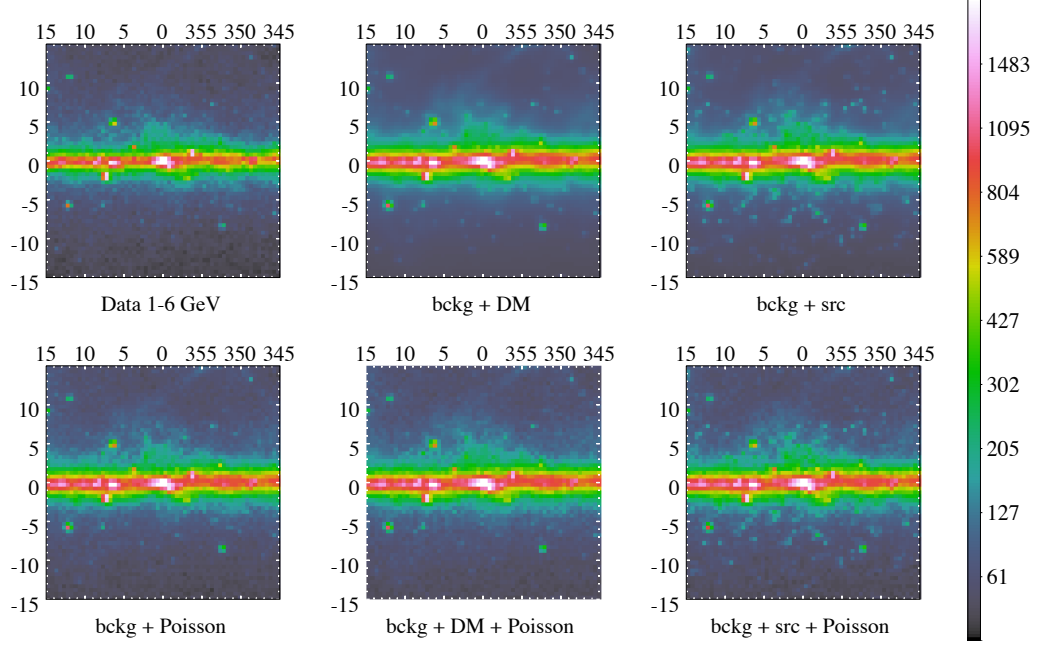


Figure 5: Images of the Galactic center in the 1-6 GeV γ ray band. Color code in counts/pixel. Upper left: real data. Upper center: a model of the GC region as result of the fit to the Training A model. In the upper right the same model but a distribution of point sources replace the gNFW template. The lower row presents the same models of the upper row but with Poissonian noise added to make them more realistic. We use the gtmcode of *Fermi* Science Tools version 11-05-00 to create the images. *Bckg* stands for background, which is emission except for the GCE. *DM* for the diffuse component and *src* the point source population below the detection threshold.

We take the following action to account for these unknowns:

The network was trained using simulations of three background models (training A, B, and C in table 1) while validating the ConvNets on models that were generated using two different background models (testing A and B in table 1). In total five background models were used. This procedure ensures that the network is sensitive to features common to all background models instead of training on characteristics of one particular background model only.

We use many initializations of the sources, and in every model simulation, their actual positions were different. Therefore the network cannot rely on fixed positions of the point sources, but only on their distribution which is determined by the parameter α (see section 3).

4 Neural network setup

This section describes the analysis setup of the convolutional neural networks that were used to make the predictions of f_{src} . The input data of the network is a (w, h, c) -tensor, with

w and h the width and height of the image and c the number of channels (colors in case of a color image). As the pixels of the image represent photon counts in the band between 1 and 6 GeV, the number of channels is one (meaning the image is monochromatic). The output of the network is a number between 0 and 1, representing the value of f_{src} . In total five networks were trained with the same network architecture, see table 2. We train one network on the full image, and other four networks trained on half of the image (left, right, top and bottom half). After training, the output values of the networks trained on half of the images can be compared to check if their predictions are in agreement. Also, the output values of these four networks can be averaged to obtain an averaged prediction of these four networks. Averaging multiple network results typically improves overall accuracy [52] (also called *ensemble learning*).

Table 2: The five different networks trained on the simulations.

Network name	Tensor size
Full image	(120, 120, 1)
Left half of image	(60, 120, 1)
Right half of image	(60, 120, 1)
Top half of image	(120, 60, 1)
Bottom half of image	(120, 60, 1)

We average the predictions of the ConvNets trained on partial images; the result is labeled 'averaged network' in what follows. The training data contains images that are generated using three different background models. The validation data consists of 60.000 images from two background models that are not part of the training data (Training A and B in table 1, we split the GC pictures evenly) and with randomized α and point source locations. Therefore, to get high accuracy with the ConvNet recovering f_{src} on the 60.000 images in the validation set, we must account for variations in the locations and fluxes of the individual unresolved sources, as well as uncertainties in the IEMs.

As a preprocessing step, we normalize the pixel values of the images to values between 0 and 1 to improve the training speed. The last layer of the network architecture (see Appendix A) has a sigmoid activation function, $\sigma(x) = (1 + e^{-x})^{-1}$. This function goes asymptotically from 0 to 1, making an $f_{\text{src}} = 1$ prediction impossible (the input of the last layer has to be infinite in this case). To improve network accuracy for very high f_{src} predictions, the output value of a prediction is normalized with a factor of $\frac{1}{\max(f_{\text{src}})}$ to map the highest predicted output to one. Here $\max(f_{\text{src}})$ is the highest predicted value of the validation data.

4.1 Network architecture

This section explains the architecture of the ConvNet used for this work. For a general high-level introduction to ConvNets, see Appendix A. The ConvNet architecture is identical for all five networks and is visualized in figure 6. The activation functions of all inner layers are ReLU functions [53]: $f(x) = x$ if $x > 0$ and $f(x) = 0$ otherwise. The last layer has a sigmoid activation function.

The loss function of the network is mean square error; the optimizer is chosen to be the Adam optimizer [54]. We train the ConvNet in two steps: first with a learning rate of 10^{-5} and then with a learning rate of 10^{-9} , both using 20 epochs [54]. We used the TensorFlow⁹

⁹<https://www.tensorflow.org>

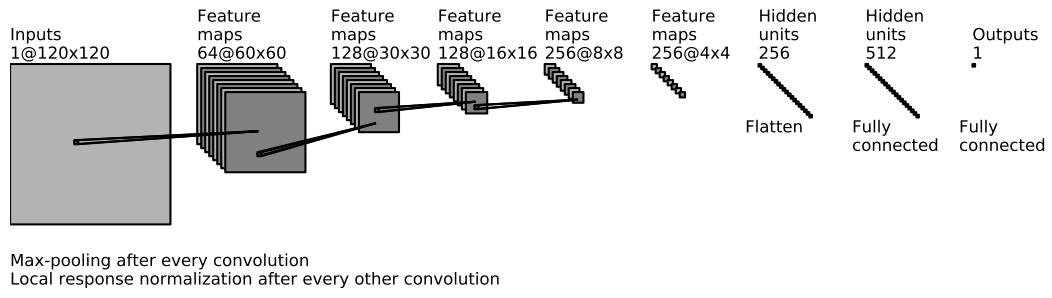


Figure 6: Visualization of the convolutional neural network. The network consists of an input layer, 5 convolutional + pooling layers, 2 fully connected layers and finally an output layer. See [Appendix A](#) for an introduction on ConvNets.

library and ran training on two Nvidia GTX1080 cards.

5 Results

5.1 Neural network results

Figure 7 shows a visualization of the different activations within the ConvNet layers. The input image on the left. Each column represents a convolutional layer, and the image passes through the network from left to right. Only four convolutional filters are shown per column for clarity. In the actual network, the number of filters varies between 64 and 256. Also, because there is a max pooling layer (see Appendix A.1.1) between the convolutional layers, the images are smaller in subsequent convolutional layers. We zoomed them to identical sizes in the visualization. The figure shows that some filters tend to filter out the diffuse background, while others seem to take the diffuse background into account only. Internally the network seems to decompose images into its diffuse and point source components. For this particular simulated sample, the difference between the ConvNet prediction and the actual value of f_{src} of the sample was 0.04.

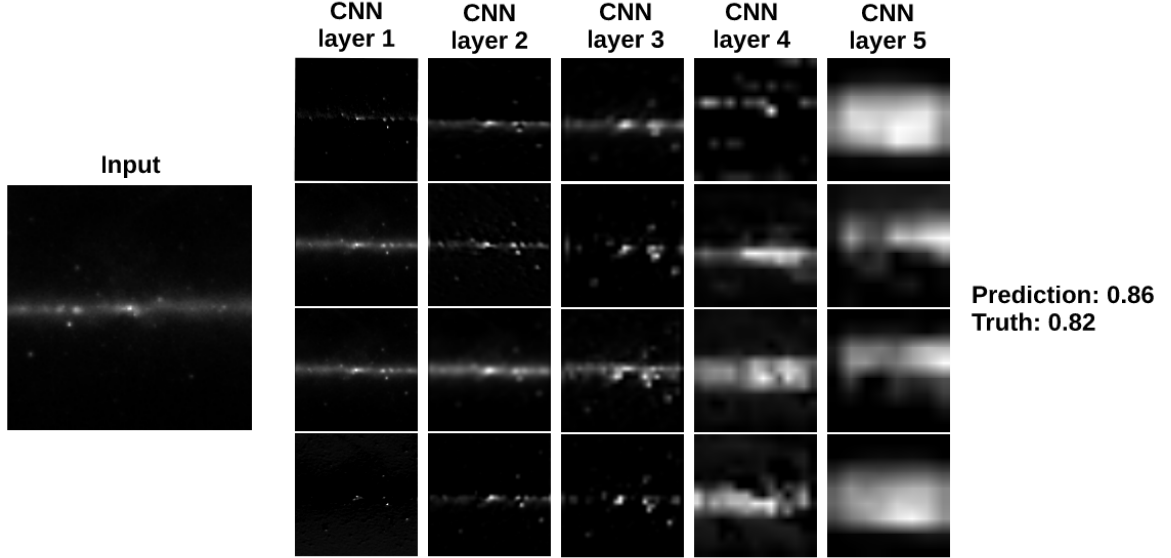


Figure 7: Example filters of different layers of the full network on a simulation. Each column of images represents a convolutional layer. The data flows from left to right (from input to prediction). Each convolutional + pooling layer accepts the image as input and outputs n smaller images, where n is the number of convolutional filters in that layer (n ranges from 64 to 256, see figure 6). Only four filters per layers are shown for clarity.

In figure 8 we show the results of the ConvNet output on the validation data for three different networks: the full network, the averaged network and the network trained on the left half of the image. Note that these are predictions on two different models that were not in the training data (Testing A and B models in table 1). We compute the accuracy of the network in 10 equally spaced bins based on this validation set. This is done because the network is more accurate for low f_{src} -predictions than for high f_{src} -predictions (see figure 8). The average network performs slightly better than the full network, meaning an ensemble of different networks achieves higher accuracy than training one network on the full image (see section 4).

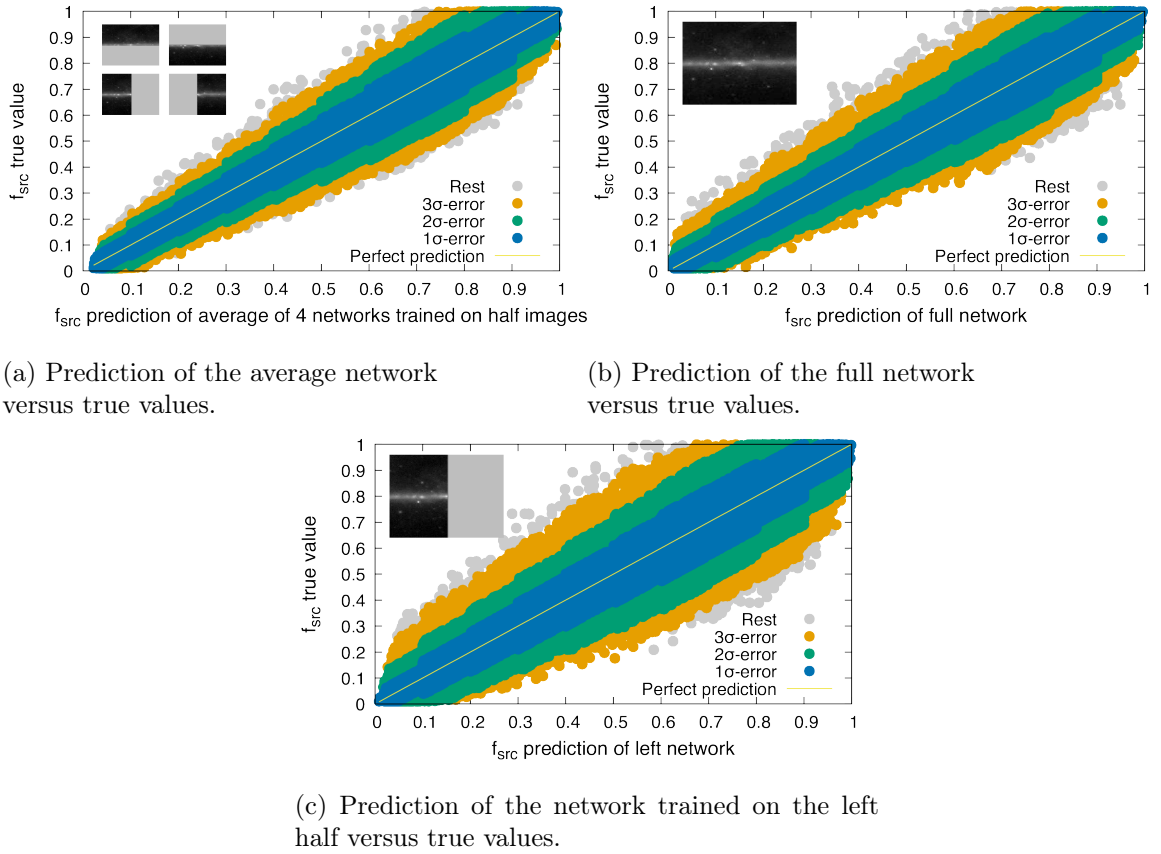


Figure 8: Network results on the validation set. The different colors represent the 1σ , 2σ and 3σ bands. The diagonal line represents a perfect prediction.

Table 3: Properties of the network trained on the left half at the predicted value on the real *Fermi* result.

Name	Value
Predicted output	0.887
1σ error	0.105
2σ error	0.210
3σ error	0.324
Maximum error	0.416

To minimize the researcher bias for the future implementation of this method only the network trained on the left half of the image was used to predict the real *Fermi*-LAT data. The output of this ConvNet is 0.887, putting the network value in the bin with the errors shown in table 3. We reserve the other half and the full picture for the follow-up work.

The predicted value for f_{src} is 0.887 ± 0.105 . This value favors an interpretation regarding point sources for the GCE. However, it is worth noticing that we must not use this ConvNet on real data as the images used for training and testing the ConvNets use the 4-years 3FGL catalog while we use about seven years of data. The newer Second Fermi Inner Galaxy catalog

(2FIG) contains about double the sources in the same region [32]. Also, we train the networks on α values between -1.05 and 1.05, but larger values are physically plausible. This will be addressed in a follow-up paper.

We use a fixed value of γ in the simulations. To see if changing γ negatively impacts the predictions few simulations were run with varying γ values. As can be seen in figure 9 this had a sizable impact on the result. For $\gamma = 1.1$ and $\gamma = 1$ the network predictions remain accurate, but for higher values of γ the network under-predicts f_{src} . Because a higher value of γ means that the point sources are closer to the GC, it becomes increasingly difficult to distinguish the emission from the point-source population from truly diffuse emission. It is therefore expected that the network under-predicts f_{src} for higher values of γ . Since the most likely value of γ we obtained in our fit is between 1 and 1.1, it is expected that the network is still accurate on the real *Fermi*-LAT data. However, the next iteration of the network will be trained on multiple values of γ to ensure that the f_{src} predictions are accurate regardless of different γ values.

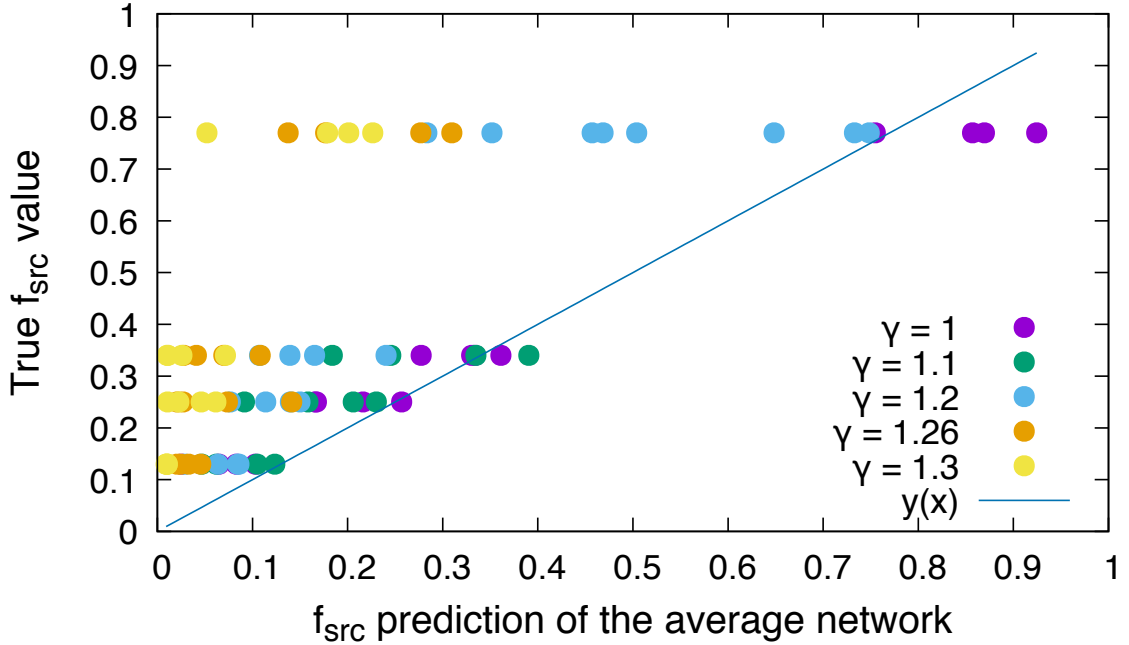


Figure 9: Predictions of the average network using different values for γ . For $\gamma = 1$ and $\gamma = 1.1$ the prediction are still accurate. However for higher values of γ the network starts to under-predict the true values of f_{src} . This shows that the network does not generalize over higher than $\gamma = 1.1$ values, but does for values between 1 and 1.1.

6 Conclusion

Training a convolutional neural network on simulated data of the *Fermi* LAT yields promising results to characterize the contribution of unresolved dim point sources to the diffuse emission. We find that the ConvNet method gives predictions beyond the classification of either 100% point sources or diffuse source: it can also predict a mixture. The error on the prediction of f_{src} is usually on the order of 10% for the averaged network. Furthermore, this method is not limited to predicting f_{src} . One can also make predictions on the properties of the source population, like α and γ , as well as the most likely background model by training networks on those particular values.

When applied to the real data, the left network prediction of the point source fraction f_{src} of the Galactic center excess is 0.887 ± 0.105 . This value disfavors the possibility that the GCE is solely caused by diffuse emission but rather by a mixture of combined radiation of unresolved point-sources and genuinely diffuse emission, in agreement with previous studies [29, 30, 32]. The present study only presents a proof of concept, and one should note that the results are biased towards higher values of f_{src} . The reason for this bias is the use of a four-years catalog of point sources, the 3FGL, whereas in this analysis we use seven years of data. More recent lists of point sources, such as the 2FIG, already contain twice as many objects in the same region. Also, to extend the range of the α parameter from ± 1.05 to at least ± 1.7 to account for best-fit values found by other studies [32, 55].

Currently, the training data of the network is a value per pixel which represents the number of photon events between 1 and 6 GeV. However, we can feed more information into the network to make more accurate predictions and allow for more generalizations. These improvements include multiple values of γ , as discussed in the previous section. Instead of a $(w, h, 1)$ -tensor representing the counts of 1-6 GeV photons, the next iteration of this network will use a (w, h, c) -tensor, where c is the number of energy bins. Each bin will contain the pixel count of some specific energy bin. With this extra dimension of the data, the network can learn between correlations in any direction of the tensor. This extra information may lead to improved performance and will allow more complex architectures. The full analysis will yield the following improvements:

- Increase sensitivity of α to realistic values from ± 1.05 to at least ± 1.7
- Generalize over multiple values of γ instead of assuming $\gamma = 1.1$
- Use the latest point source catalog instead of the 3FGL
- Use multiple energy bins as channels instead of one channel

Appendices

A ConvNets

Recently the field of deep learning has received a lot of attention because of the predictive power of deep neural networks. These networks can be used to detect objects in images [43], create text-to-speech algorithms [56] and many more applications. This novel approach utilizes the many advances in machine learning of the past years and can lead to better predictions using fewer assumptions, as ConvNets typically require the raw data as input. Using DNNs over conventional data analysis methods has up- and downsides:

- DNNs work with raw data and learn to recognize correlations in the data automatically. This has two advantages: 1) there is no need to manually prune the data (manually de-noise, reparametrize or otherwise reduce the dimensionality of the data) and 2) the raw data contains all possible information. Pruned data does not. A conventional method that uses preprocessed data cannot access all the information enclosed in the raw data. The quality of the pruned data is entirely dependent on the human understanding of the data and it might happen one unknowingly removes correlations when preprocessing data.
- DNNs are very general. The network architecture changes from problem to problem, but the layer types and methods can be used in many different problem areas like analyzing image or video data, regression, speech analysis and many more domains. This property of DNNs means the advances in one area of research (for example face detection) also benefit all the other research areas.
- DNNs can generalize over randomness. When a DNN needs to learn magnitudes of stars in a 2D image, the randomized location of the star carries no information. By using convolutional layers and enough training samples, the network can learn to 'ignore' the location of the star. This generalization is an important feature in the analysis in this work: the unresolved point sources have a random location in our simulations.
- DNNs are hard to train. The training process tunes many hyperparameters to specific values for the network to make accurate predictions. These hyperparameters need to be set before training and require knowledge of the problem, the network, and trial-and-error.
- DNNs require a lot of data and processing power. This is one of the main reasons DNNs are only becoming popular in recent years. The hardware required to train networks that can make accurate predictions on real-world data is only available since the advent of the GPU. Also, deep neural networks typically require a lot of training examples to train well.

A.1 ConvNet data pipeline

In general, a ConvNet takes an N -dimensional input, transforms it using different layers, and produces an M -dimensional output. This can be a prediction that a specific input belongs to a particular class (e.g., object detection) or a prediction of a regression problem (e.g., this research). The input of a computer vision problem is typically an image. The image can be

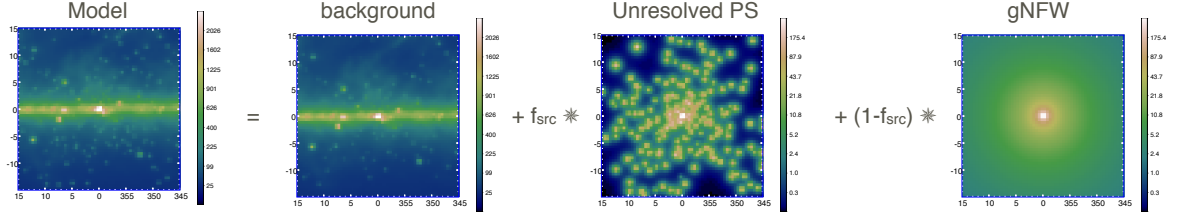


Figure 10: Visualization of the model construction. The model (that is used as training data for the ConvNets) is generated by adding a fraction f_{src} of point sources and a fraction of $1 - f_{src}$ of the diffuse source to the background. The sum of the diffuse and point source fraction that is added on top on the background always equals the GC excess.

represented by a (w, h, c) -tensor, where w and h represent the width and height of the network and c the number of channels. For instance: a 120×120 color image has $120 \cdot 120 \cdot 3 = 43.200$ different values embedded into it. In this analysis, the number of channels is one, and it represents the photon count in the energy bin 1-6 GeV. The input is the first layer of a neural network and in the example consists of 43.200 neurons. When an image is fed into the network, each input neuron consists of one color of one-pixel value. In left-hand side of figure 10 we present an example of an image for training. It is composed of a background emission, and the GCE, whose granularity is controlled by the *fraction of point sources* f_{src} parameter, which varies between 0 and 1. A value of $f_{src} = 0$ means the GCE is composed of only a diffuse source, and $f_{src} = 1$ means the GCE is composed of only point sources. With ConvNets we transform such images into predictions of the parameter f_{src} .

In ConvNets, between the input and output layers, there are so-called *hidden layers*. In this work we use the following kinds of hidden layers: fully connected, convolutional, max pooling and local response normalization. The differences of these layers will be discussed in the next section.

A.1.1 Layer Architectures

A fully connected layer connects all neurons of a layer to the next one. A weight w is assigned to each connection (see figure 11). The value $n_{y,j}$ of a particular neuron j in layer y , is defined as $n_{y,j} = f(\sum_i n_{x,i} \cdot w_{x,i;y,j} + \text{bias})$, where $f(x)$ is called the activation function. In this research the internal layers have a Rectified Linear Unit (ReLU) activation function ($f(x > 0) = x$ and $f(x \leq 0) = 0$). The output layer has a sigmoid activation function, to map all input values to the range (0,1), which represents the output range of f_{src} parameter.

During training, an image is fed into the network, and an f_{src} output value is predicted. This output can be compared to the actual f_{src} used to create the image. The error is propagated back into the network which *adjusts* all w and biases [57]. Before training all weights and biases are set to normal distributed random numbers (the network initialization).

A DNN with only fully connected layers has many parameters (weights and biases) to train. For instance: a 120×120 color image has 43.200 input neurons. If the neural network has one hidden layer with 100 neurons and one output neuron, the number of weights in the ConvNet is $43.200 \cdot 100 + 100 \cdot 1 = 4.320.100$. This makes fully connected networks very hard to train.

A convolutional layer is a layer that has fewer weights than a fully connected layer by utilizing the fact that there is a correlation between neighboring pixels on an image [40]. Max

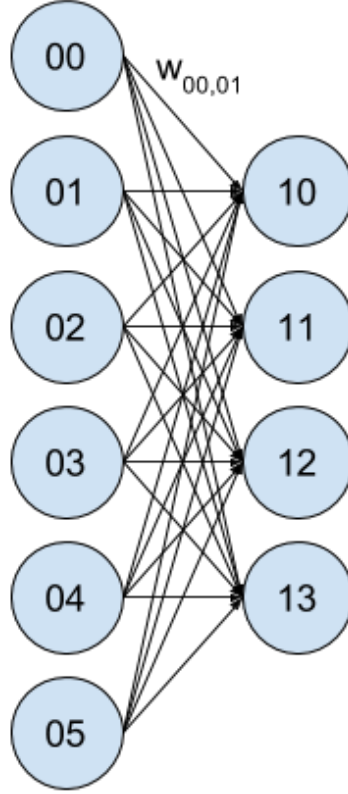


Figure 11: Schematic view of a fully connected layer with 6 input and 4 output neurons. The 24 lines represent the weights of the connections between the the neurons from the two layers.

pooling layers take a patch of an image and return the maximum of this value [58]. A max pooling layer is used to reduce the dimensions of an image during network propagation. After multiple convolutional and pooling layers, fully connected layers are added to make a final prediction of the image. A graphical representation of a typical convolutional neural network can be found in figure 2 of [40].

A.1.2 Choice of activation function

The choice of the activation functions of the neurons is crucial, as this introduces a non-linearity in the network and the right activation function optimizes the speed and predictive power of the network. If the activation function would be linear, for example, $f(x) = x$, the neural network would just be a linear combination of linear functions. If this is the case, it is impossible to make predictions about non-linear problems. Popular choices are the sigmoid and hyperbolic tangent (tanh) function because they are bound between 0 and +1 for the sigmoid function and -1 and +1 for the tanh function. However, they suffer from the so-called vanishing gradient problem [59]: the gradient at significant positive or negative values go to zero. As the gradient of the activation function is a used in the backpropagation algorithm

to update the weights, the network effectively stops learning when the gradient is (almost) zero.

The ReLU activation function is currently a popular choice, because it has no vanishing gradient and is a non-linear function[53]. On top of that, the derivative of the ReLU function is computationally trivial, while calculating the derivative of a sigmoid or tanh function can be computationally expensive. However, the ReLU function has no gradient for negative values, leading to so-called 'dying ReLUs': if the update of a weight makes the output of the ReLU negative, the particular neuron is effectively dead as there is no gradient anymore. This means the weights will not be updated anymore during training. There are many more activation functions that can be chosen and each of them has advantages and disadvantages. In this work, the ReLU activation function was chosen because of its speed and non-vanishing gradient.

A.2 Training a neural network

A neural network has many weights in it, which are initialized randomly according to some distribution (typically normal distributed). These weights need to be changed from the random initialization to particular values, to predict the training dataset accurately. This is done during the training phase of the network. During training, the network receives an image and calculates an (incorrect) output. The error is propagated back into the network, and the weights are set so the error of the prediction is less when the image would go through the network again. There are many backpropagation algorithms, of which the Adam optimizer is a popular one and the one used in this research [54].

A typical difficulty of training neural networks is overfitting. In this case, the network can predict every example in the training data perfectly, while it fails to generalize over the important features. Instead of learning those, it 'memorized' the training data and cannot be used for other cases. To determine whether a network is overfitted to the training data, one can set aside a part of the dataset and not use it during training. After the network is trained, the accuracy of the network on this validation set is calculated. It is an excellent indication that the network has overfitted if the training set accuracy is much higher than the validation set accuracy [60]. There are a number of methods to prevent overfitting, such as regularization [61] and dropout layers [62]. In this research, the convolutional layers have an L2 regularizer to penalize large weights. Adding dropout has been considered as well, but only using L2 regularizers was enough to prevent overfitting and adding a dropout layer did not improve the network accuracy.

To force the network not to overfit to particular values that are not known (the actual positions of the point sources, α , or the specific background model used for example), many images are generated that span the possible space of values. The network has to generalize over these values to achieve reasonable accuracy. This means many realizations of the simulations are needed. As long as the real data is somewhere inside the training box spanned by f_{src} , randomized locations of unresolved point source population and the background model, and the ConvNet generalizes over all these values, the network output is reliable.

Acknowledgements

The *Fermi* LAT Collaboration acknowledges generous on-going support from a number of agencies and institutes that have supported both the development and the operation of the LAT as well as scientific data analysis. These include the National Aeronautics and Space Administration and the Department of Energy in the United States; the Commissariat à l’Energie Atomique and the Centre National de la Recherche Scientifique/Institut National de Physique Nucléaire et de Physique des Particules in France; the Agenzia Spaziale Italiana and the Istituto Nazionale di Fisica Nucleare in Italy; the Ministry of Education, Culture, Sports, Science and Technology (MEXT), High Energy Accelerator Research Organization (KEK), and Japan Aerospace Exploration Agency (JAXA) in Japan; and the K. A. Wallenberg Foundation, the Swedish Research Council, and the Swedish National Space Board in Sweden. Additional support for science analysis during the operations phase is gratefully acknowledged from the Istituto Nazionale di Astrofisica in Italy and the Centre National d’Etudes Spatiales in France.

The authors thanks Mattia di Mauro, Manuel Meyer, Gabriela Zaharijas, Christoph Weniger and Tom Heskes for fruitful comments on the manuscript. R. RdA, is supported by the Ramón y Cajal program of the Spanish MICINN and also thanks the support of the Spanish MICINN’s Consolider-Ingenio 2010 Programme under the grant MULTIDARK CSD2209-00064, the Invisibles European ITN project (FP7-PEOPLE-2011-ITN, PITN-GA-2011-289442-INVISIBLES, the “SOM Sabor y origen de la Materia” (FPA2011-29678), the “Fenomenologia y Cosmologia de la Fisica mas alla del Modelo Estandar e Implicaciones Experimentales en la era del LHC” (FPA2010-17747) MEC projects and the Spanish MINECO Centro de Excelencia Severo Ochoa del IFIC program under grant SEV-2014-0398. The work of GAGV was supported by Programa FONDECYT Postdoctorado under grant 3160153.

SC and LH acknowledge the support within the "idark" program of the Netherlands eScience Center (NLeSC).

References

- [1] W. B. Atwood, A. A. Abdo, M. Ackermann, W. Althouse, B. Anderson, M. Axelsson, L. Baldini, J. Ballet, D. L. Band, G. Barbiellini, and et al., “The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission,” *Astrophysical Journal*, vol. 697, pp. 1071–1102, June 2009.
- [2] M. Ackermann *et al.*, “The Fermi Large Area Telescope On Orbit: Event Classification, Instrument Response Functions, and Calibration,” *Astrophys. J. Suppl.*, vol. 203, p. 4, 2012.
- [3] I. V. Moskalenko and A. W. Strong, “Production and propagation of cosmic ray positrons and electrons,” *Astrophys. J.*, vol. 493, pp. 694–707, 1998.
- [4] A. W. Strong, I. V. Moskalenko, and O. Reimer, “Diffuse continuum gamma-rays from the galaxy,” *Astrophys. J.*, vol. 537, pp. 763–784, 2000. [Erratum: *Astrophys. J.* 541, 1109 (2000)].
- [5] J. R. Mattox, D. L. Bertsch, J. Chiang, B. L. Dingus, S. W. Digel, J. A. Esposito, J. M. Fierro, R. C. Hartman, S. D. Hunter, G. Kanbach, D. A. Kniffen, Y. C. Lin, D. J. Macomb, H. A. Mayer-Hasselwander, P. F. Michelson, C. von Montigny, R. Mukherjee, P. L. Nolan, P. V. Ramanamurthy, E. Schneid, P. Sreekumar, D. J. Thompson, and T. D. Willis, “The Likelihood Analysis of EGRET Data,” *Astrophys. J.*, vol. 461, p. 396, Apr. 1996.
- [6] M. Ackermann *et al.*, “Fermi-LAT Observations of the Diffuse Gamma-Ray Emission: Implications for Cosmic Rays and the Interstellar Medium,” *Astrophys. J.*, vol. 750, p. 3, 2012.

- [7] L. Goodenough and D. Hooper, “Possible Evidence For Dark Matter Annihilation In The Inner Milky Way From The Fermi Gamma Ray Space Telescope,” *ArXiv e-prints*, Oct. 2009.
- [8] V. Vitale and A. Morselli, “Indirect Search for Dark Matter from the center of the Milky Way with the Fermi-Large Area Telescope,” in *Fermi gamma-ray space telescope. Proceedings, 2nd Fermi Symposium, Washington, USA, November 2-5, 2009*, 2009.
- [9] D. Hooper and L. Goodenough, “Dark Matter Annihilation in The Galactic Center As Seen by the Fermi Gamma Ray Space Telescope,” *Phys. Lett.*, vol. B697, pp. 412–428, 2011.
- [10] C. Gordon and O. Macias, “Dark Matter and Pulsar Model Constraints from Galactic Center Fermi-LAT Gamma Ray Observations,” *Phys. Rev.*, vol. D88, no. 8, p. 083521, 2013. [Erratum: *Phys. Rev.*D89,no.4,049901(2014)].
- [11] D. Hooper and T. Linden, “On The Origin Of The Gamma Rays From The Galactic Center,” *Phys.Rev.*, vol. D84, p. 123005, 2011.
- [12] T. Daylan, D. P. Finkbeiner, D. Hooper, T. Linden, S. K. N. Portillo, N. L. Rodd, and T. R. Slatyer, “The characterization of the gamma-ray signal from the central Milky Way: A case for annihilating dark matter,” *Phys. Dark Univ.*, vol. 12, pp. 1–23, 2016.
- [13] A. Boyarsky, D. Malyshev, and O. Ruchayskiy, “A comment on the emission from the Galactic Center as seen by the Fermi telescope,” *Physics Letters B*, vol. 705, pp. 165–169, Nov. 2011.
- [14] F. Calore, I. Cholis, and C. Weniger, “Background model systematics for the Fermi GeV excess,” *JCAP*, vol. 3, p. 038, Mar. 2015.
- [15] K. N. Abazajian, N. Canac, S. Horiuchi, and M. Kaplinghat, “Astrophysical and Dark Matter Interpretations of Extended Gamma-Ray Emission from the Galactic Center,” *Phys. Rev.*, vol. D90, no. 2, p. 023526, 2014.
- [16] B. Zhou, Y.-F. Liang, X. Huang, X. Li, Y.-Z. Fan, L. Feng, and J. Chang, “GeV excess in the Milky Way: The role of diffuse galactic gamma-ray emission templates,” *Phys. Rev.*, vol. D91, no. 12, p. 123010, 2015.
- [17] **Fermi/LAT** Collaboration, “Fermi-LAT Observations of High-Energy Gamma-Ray Emission toward the Galactic Center,” *ApJ*, vol. 819, p. 44, Mar. 2016.
- [18] M. Ackermann *et al.*, “The Fermi Galactic Center GeV Excess and Implications for Dark Matter,” *Astrophys. J.*, vol. 840, no. 1, p. 43, 2017.
- [19] A. Achterberg, S. Amoroso, S. Caron, L. Hendriks, R. Ruiz de Austri, and C. Weniger, “A description of the Galactic Center excess in the Minimal Supersymmetric Standard Model,” *JCAP*, vol. 1508, no. 08, p. 006, 2015.
- [20] N. Mirabal, “Dark matter vs. Pulsars: Catching the impostor,” *Mon. Not. Roy. Astron. Soc.*, vol. 436, p. 2461, 2013.
- [21] C. Karwin, S. Murgia, T. M. P. Tait, T. A. Porter, and P. Tanedo, “Dark matter interpretation of the fermi-lat observation toward the galactic center,” *Phys. Rev. D*, vol. 95, p. 103005, May 2017.
- [22] A. McCann, “A stacked analysis of 115 pulsars observed by the Fermi LAT,” *Astrophys. J.*, vol. 804, no. 2, p. 86, 2015.
- [23] R. M. O’Leary, M. D. Kistler, M. Kerr, and J. Dexter, “Young Pulsars and the Galactic Center GeV Gamma-ray Excess,” *ArXiv:1504.02477*, Apr. 2015.
- [24] Q. Yuan and B. Zhang, “Millisecond pulsar interpretation of the Galactic center gamma-ray excess,” *Journal of High Energy Astrophysics*, vol. 3, pp. 1–8, Sept. 2014.
- [25] J. Petrović, P. D. Serpico, and G. Zaharijas, “Millisecond pulsars and the Galactic Center gamma-ray excess: the importance of luminosity function and secondary emission,” *JCAP*, vol. 1502, no. 02, p. 023, 2015.

- [26] I. Cholis, D. Hooper, and T. Linden, “Challenges in Explaining the Galactic Center Gamma-Ray Excess with Millisecond Pulsars,” *JCAP*, vol. 1506, no. 06, p. 043, 2015.
- [27] O. Macias, C. Gordon, R. M. Crocker, B. Coleman, D. Paterson, S. Horiuchi, and M. Pohl, “Discovery of Gamma-Ray Emission from the X-shaped Bulge of the Milky Way,” 2016.
- [28] J.-P. Macquart and N. Kanekar, “On Detecting Millisecond Pulsars at the Galactic Center,” *Astrophys. J.*, vol. 805, no. 2, p. 172, 2015.
- [29] S. K. Lee, M. Lisanti, B. R. Safdi, T. R. Slatyer, and W. Xue, “Evidence for Unresolved γ -Ray Point Sources in the Inner Galaxy,” *Physical Review Letters*, vol. 116, p. 051103, Feb. 2016.
- [30] R. Bartels, S. Krishnamurthy, and C. Weniger, “Strong Support for the Millisecond Pulsar Origin of the Galactic Center GeV Excess,” *Physical Review Letters*, vol. 116, p. 051102, Feb. 2016.
- [31] E. Storm, C. Weniger, and F. Calore, “SkyFACT: High-dimensional modeling of gamma-ray emission with adaptive templates and penalized likelihoods,” *JCAP*, vol. 1708, no. 08, p. 022, 2017.
- [32] M. Ajello *et al.*, “Characterizing the population of pulsars in the Galactic bulge with the *Fermi* Large Area Telescope,” *Submitted to: Astrophys. J.*, 2017.
- [33] D. Hooper and G. Mohlabeng, “The Gamma-Ray Luminosity Function of Millisecond Pulsars and Implications for the GeV Excess,” *JCAP*, vol. 1603, no. 03, p. 049, 2016.
- [34] D. Hooper and T. Linden, “The Gamma-Ray Pulsar Population of Globular Clusters: Implications for the GeV Excess,” *JCAP*, vol. 1608, no. 08, p. 018, 2016.
- [35] D. Haggard, C. Heinke, D. Hooper, and T. Linden, “Low Mass X-Ray Binaries in the Inner Galaxy: Implications for Millisecond Pulsars and the GeV Excess,” *JCAP*, vol. 1705, no. 05, p. 056, 2017.
- [36] C. Eckner *et al.*, “Millisecond pulsar origin of the Galactic center excess and extended gamma-ray emission from Andromeda - a closer look,” 2017.
- [37] R. Bartels, E. Storm, C. Weniger, and F. Calore, “The *Fermi*-LAT GeV Excess Traces Stellar Mass in the Galactic Bulge,” 2017.
- [38] H. Ploeg, C. Gordon, R. Crocker, and O. Macias, “Consistency Between the Luminosity Function of Resolved Millisecond Pulsars and the Galactic Center Excess,” *JCAP*, vol. 2017, no. 08, p. 015, 2017.
- [39] P. Agrawal and L. Randall, “Point Sources from Dissipative Dark Matter,” 2017.
- [40] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [41] D. George, H. Shen, and E. A. Huerta, “Deep Transfer Learning: A new deep learning glitch classification method for advanced LIGO,” 2017.
- [42] Y. D. Hezaveh, L. Perreault Levasseur, and P. J. Marshall, “Fast Automated Analysis of Strong Gravitational Lenses with Convolutional Neural Networks,” *Nature*, vol. 548, pp. 555–557, 2017.
- [43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.
- [44] F. de Palma, T. J. Brandt, J. Hewitt, G. Johannesson, and L. Tibaldo, “Systematically characterizing regions of the First *Fermi*-LAT SNR Catalog,” *PoS*, vol. ICRC2015, p. 814, 2016.
- [45] F. Acero *et al.*, “Development of the Model of Galactic Interstellar Emission for Standard Point-Source Analysis of *Fermi* Large Area Telescope Data,” *Astrophys. J. Suppl.*, vol. 223, no. 2, p. 26, 2016.

- [46] F. Acero *et al.*, “Fermi Large Area Telescope Third Source Catalog,” *Astrophys. J. Suppl.*, vol. 218, no. 2, p. 23, 2015.
- [47] J. F. Navarro, C. S. Frenk, and S. D. M. White, “A Universal Density Profile from Hierarchical Clustering,” *ApJ*, vol. 490, pp. 493–508, Dec. 1997.
- [48] J. F. Navarro, C. S. Frenk, and S. D. White, “The Structure of cold dark matter halos,” *Astrophys. J.*, vol. 462, pp. 563–575, 1996.
- [49] A. A. Abdo, M. Ackermann, M. Ajello, A. Allafort, E. Antolini, W. B. Atwood, M. Axelsson, L. Baldini, J. Ballet, G. Barbiellini, and *et al.*, “Fermi Large Area Telescope First Source Catalog,” *Astrophysical Journal, Supplement*, vol. 188, pp. 405–436, June 2010.
- [50] D. Lorimer, A. Faulkner, A. Lyne, *et al.*, “The Parkes multibeam pulsar survey: VI. Discovery and timing of 142 pulsars and a Galactic population analysis,” *Mon. Not. Roy. Astron. Soc.*, vol. 372, pp. 777–800, 2006.
- [51] G. L. Case and D. Bhattacharya, “A new sigma-d relation and its application to the galactic supernova remnant distribution,” *Astrophys. J.*, vol. 504, p. 761, 1998.
- [52] R. Polikar, “Ensemble based systems in decision making,” *IEEE Circuits and Systems Magazine*, vol. 6, no. 3, p. 021, 2006.
- [53] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (J. Fürnkranz and T. Joachims, eds.), pp. 807–814, Omnipress, 2010.
- [54] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *CoRR*, vol. abs/1412.6980, 2014.
- [55] R. Bartels, D. Hooper, T. Linden, S. Mishra-Sharma, N. L. Rodd, B. R. Safdi, and T. R. Slatyer, “Comment on “Characterizing the population of pulsars in the Galactic bulge with the Fermi Large Area Telescope” [arXiv:1705.00009v1],” 2017.
- [56] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. W. Senior, and K. Kavukcuoglu, “Wavenet: A generative model for raw audio,” *CoRR*, vol. abs/1609.03499, 2016.
- [57] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, “Efficient backprop,” in *Neural Networks: Tricks of the Trade, This Book is an Outgrowth of a 1996 NIPS Workshop*, (London, UK, UK), pp. 9–50, Springer-Verlag, 1998.
- [58] M. D. Zeiler and R. Fergus, *Visualizing and Understanding Convolutional Networks*, pp. 818–833. Cham: Springer International Publishing, 2014.
- [59] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, “Gradient flow in recurrent nets: the difficulty of learning long-term dependencies,” 2001.
- [60] R. Caruana, S. Lawrence, and C. L. Giles, “Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping,” in *Advances in Neural Information Processing Systems 13* (T. K. Leen, T. G. Dietterich, and V. Tresp, eds.), pp. 402–408, MIT Press, 2001.
- [61] A. Y. Ng, “Feature selection, l1 vs. l2 regularization, and rotational invariance,” in *Proceedings of the Twenty-first International Conference on Machine Learning, ICML ’04*, (New York, NY, USA), pp. 78–, ACM, 2004.
- [62] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, Jan. 2014.